山形大学紀要（工学）第22巻 第2号 平成5年1月
Bull. Yamagata Univ. (Eng.) Vol.22 No. 2 Jan. 1993

Polymorphism of Repressibility of σ^{70}-promoter-like Sequences in Escherichia coli and its Bacteriophages.

Shigehiko KANAYA and Yoshihiro KUDO

Department of Electrical and Information Engineering, Faculty of Engineering.

(*Received September* 1, 1992)

Abstracts

The present paper compares, among Escherichia coli and 9 bacteriophages, repressibilities of occurrences of σ^{70}-promoter consensus sequences in coding regions by means of an index (Ks) derived from a χ^2-statistic and principal components analysis(PCA). Results of PCA represent a subtle distinction of the gene distributions among E. coli and its bacteriphages as follows: 〔1〕 the occurrences of the promoter-like sequences for E. coli, λ and P4 are repressed and degree of the repression for E. coli is the highest of the three, and 〔2〕 the occurrences of the promoter-like sequences for filamentous phages (M13, fd, fl, and IKe), T7, G4 and ΦX174 are promotive and the occurrences for filamentaous phages are more promotive than those for T7, G4 and ΦX174. These results suggest that the repressibilty of the promoter-like sequences in coding regions is one of the most important factors to effectively coordinate a web of gene control circuits.

1. Introduction

There is a promoter consensus sequence consisting of two 6-base ones, 5'TTGACA 3' (the -35region) and 5'TATAAT 3' (the -10 region), upstream of transcription initiation sites of Escherichia coli and its bacteriphages genes transcribed by E. coli σ^{70}-RNA polymerase[1-4].

Investigating of repressibility of transcription initiation in coding regions in E. coli and its bacteriophages genomes, we have developed a new statistical index, Ks, derived from χ^2-statistic and reported on repressibility of occurrences of the promoter-like sequences in coding regions[5-8].

The present paper compares, among E. coli and its bacteriophages, repressibilities of occurrences of the promoter-like sequences in coding regions by means of the Ks index and principal components analysis (PCA), and evaluates a distinction among E. coli, temperate phages (P4 and λ), filamentous phages (M13, fd, fl and IKe) and others (T7, ΦX174 and G4).

2. Methods

Details of some properties of the statistical index, Ks, were reported elsewhere[6]. In a coding region between the i-th and t-th bases (for potential initiation and termination sites, respectively), all pairs of two 6-base sequences, (5'b_{16} to b_{26} 3') and (5'b_{21} to b_{26} 3'), separated by s-base spacer, are compared with the consensus sequence (5' TTGACA 3' 5' TATAAT 3'), base by base, number k's are counted up when b_{11}, b_{12},

b_{21}, b_{23} or b_{26} is T, b_{14}, b_{16}, b_{22}, b_{24} or b_{25} is A, b_{13} is G or b_{15} is C, and fs(k)'s are counted up, where fs (k) is a frequency of k-base matched sequences. The statistical index Ks represents a difference of two global gravities of negative and positive χ -values as follow :

$$Ks=\frac{\sum_k\{kx_s(k)\}|_{x_s(k)>0}}{\sum_k x_s(k)\ |_{x_s(k)>0}}-\frac{\sum_k\{kx_s(k)\}|_{x_s(k)<0}}{\sum_k x_s(k)\ |_{x_s(k)<0}}$$

where $\chi_s(k)=\{fs(k)-Ts(k)\}/\sqrt{Ts(k)}$.
Here, Ts(k) is the theoretical frequency and is simply approximated as follows,

$$Ts(k)=(t-i+12+s)_{12}C_k(1/4)^k(3/4)^{12-k}$$

because the amounts of the four kinds of bases are almost equal to each other in E. coli genome. The index, Ks, reflects the degree of repression of the promoter-like sequences in coding units : The larger absolute value of a negative Ks of a unit is, the more strongly the promoter-like sequences are repressed in the unit.

3. Results

Table 1 shows E. coli and its bacteriophages genomic coding regions whose DNA sequences are extracted from GenBank (Release69.0) and analyzed in this study. The numbers of genomic genes analyzed in E. coli and its bacteriophages ΦX174, G 4, M13, fd, fl, IKe, P 4, T 7 and λ are 838(corresponding to about 23% of the number of genomic genes estimated[9]) and 11, 11, 10, 10, 10, 10, 13, 56 and 61, respectively.

Figure 1 shows distributions of percent gene frequencies of10 classes defined with the signs of the indices Ks' and five levels of statistical significance by χ^2- statistic($p<0.0025$, $0.0025\leqq p<0.005$, $0.005\leqq p<0.025$, $0.025\leqq p<0.05$, and $0.05\leqq p$) for transcribed (TS) and non-transcribed (NTS) strands. For all the spacer types on the both strands (TS and NTS), about 82% of E. coli genomic genes analyzed (cf.45% in the theoretical frequency)are of negative Ks' and about a half (42%) of the 82% belong to a class, $p<0.0025$. The indices Ks' of most of genes for temperate phages (P 4[10] and λ[11]) are negative, while those for other phages (ΦX174, G 4, M13, fd, fl, IKe and T 7) are positive.

PCA was applied to data sets comprising 70 kinds of the percent gene frequencies(<E. coli and 9 kinds of bacteriophages>×< 7 types of spacers>) represented as 10-dimensional vectors on the each strand. Figures 2 and 3 show plots of factor loadings and principal components by the first two components. The percent variances for the first and second principal components are 56.6% and 32.0% for TS, and 62.5% and 28.8% for NTS, respectively. On the both strands, the first and second factors mainly express negative correlations between all the five classes for Ks < 0 (N 1 to N 5 in Figure 2; negative) and the class of $p<0.0025$ for Ks> 0 (P 5; positive), and between all the five classes for Ks < 0 (N 1 to N 5; positive)and the class of $p\geqq 0.05$ for Ks> 0 (P 5; negative), respectively.

For E. coli, λ, P4 and T7, variations of the PC projections of all spacer types are smaller than those of the other phages. E. coli and filamentous phages (IKe[12], M13[13], fd[14] and fl[14]) have negative and positive PC1's, respectively. The two temperate phages are the nearest to E. coli in the PC projections.

Table 1. Coding Regions Analyzed

Table 1 (a) Escherichia coli

	Gene	Length (bases)	GenBank Locus
1.	thr (Leader)	66	ECOTHR
2.	thrA	2463	ECOTHR
3.	thrB	930	ECOTHR
4.	thrC	1287	ECOTHR
5.	dnaJ	1131	ECODNAJ, ECODNAJK
6.	ant	1089	ECOANTAPA
7.	ORF (28KD)	789	ECORPST
8.	rpsT	264	ECORPST
9.	protein X	939	ECORPSTA
10.	lspA	495	ECOLSP, ECOLSPA
11.	ORF1	450	ECOLSPDAP
12.	ORF2	951	ECOLSPDAP
13.	ORF3	915	ECOLSPDAP
14.	carA	1149	ECOCARAB
15.	carB	3222	ECOCARAB, ECOCARB
16.	kefC	1863	ECOKEFC
17.	apaH	843	ECOAPAH
18.	apaG	378	ECOAPAH
19.	ksgA	822	ECOAPAH
20.	polB	2307	ECOPOLB
21.	polB	2313	ECOPOLBDA
22.	polB	2352	ECOPOLD
23.	araD	696	ECOARAABD
24.	araD	696	ECOARADG, ECOPOLB
25.	araA	1503	ECOARAABD
26.	araB	1701	ECOARAABD
27.	araC	879	ECOARACB
28.	araC	879	ECOARACK
29.	leu (leader)	87	ECO2MIN
30.	leuO	1050	ECO2MIN
31.	ilvI	1701	ECO2MIN
32.	ilvI	1701	ECO1LVIH
33.	ilvH	492	ECO2MIN
34.	ilvH	492	ECO1LVIH
35.	shl (leader)	87	ECO2MIN, ECOSHL
36.	shl (fruR)	1005	ECO2MIN, ECOSHL, ECOFRURG
37.	ORFC	459	ECO2MIN
38.	ORFB	942	ECO2MIN
39.	ORFA	367	ECO2MIN
40.	pbp3	1767	ECO2MIN
41.	murE	1488	ECO2MIN
42.	murF	1359	ECO2MIN
43.	mraY	1083	ECO2MIN
44.	murD	1317	ECO2MIN
45.	ftsW	1245	ECO2MIN, ECOFTSW
46.	murG	1044	ECO2MIN
47.	murG	1065	ECOMURG, ECOMURGC
48.	murC	1476	ECO2MIN
49.	murC	1473	ECOMURGC
50.	ddl	921	ECO2MIN, ECODDLFTS, ECOFTSQA
51.	ftsQ	831	ECO2MIN, ECOFTSQA, ECOFTSQAZ
52.	ftsA	1263	ECO2MIN, ECOFTSQA, ECOFTSQAZ
53.	ftsZ	1152	ECO2MIN, ECOFTSQAB
54.	envA	918	ECO2MIN, ECOENVAA
55.	ORF (gene X)	444	ECO2MIN, ECOSECA
56.	secA	2706	ECO2MIN, ECOSECA

ID	Gene	Length (bases)	GenBank Locus
57.	mutT	390	ECO2MIN, ECOMUTT
58.	guaC	1041	ECOGUAC
59.	aroP	1368	ECOAROP
60.	ORF (gene A)	711	ECOACE
61.	aceE	2661	ECOACE
62.	aceF	1893	ECOACE
63.	lpd	1425	ECOACE
64.	speD	795	ECOSPDE
65.	speE	867	ECOSPDE
66.	ORF	348	ECOSPDE
67.	fhuA	2244	ECOFHUACD
68.	fhuC	798	ECOFHUACD
69.	fhuC	798	ECOFHUACE
70.	fhuD	891	ECOFHUACD
71.	fhuD	891	ECOFHUACE
72.	fhuB	1980	ECOFHUB
73.	fhuB (alt.)	1905	ECOFHUB
74.	dapD	825	ECODAPD
75.	rpsB	723	ECORPSB
76.	rpsB	726	ECORPSBTS
77.	tsf	849	ECORPSB
78.	tsf	852	ECORPSBTS
79.	cdsA	750	ECOCDS
80.	firA	1026	ECOFIRA
81.	lpxA	789	ECOLPXA
82.	lpxB	1149	ECOLPXA
83.	dnaE	3483	ECOLPXA
84.	drpA	1551	ECODRPA
85.	aspV (tRNA)	77	ECOASPV
86.	dnaQ (=mutD)	732	ECOMUTD, ECORNHQ
87.	rnh	468	ECOMUTD, ECORNH, ECORNHQ
88.	pepD	1458	ECOPEPD
89.	gpt	459	ECOGPT, ECOGPTA, ECOGPTC
90.	ORF (49kD)	1302	ECOGPTA
91.	phoE	1056	ECOPHOE
92.	phoE	1053	ECOPHOF
93.	proB	1104	ECOPHOEA
94.	proA	1224	ECOPHOEA
95.	hns	414	ECOHNS
96.	argF	1005	ECOARGF
97.	mvrA	807	ECOMVRA
98.	betA	1668	ECOBET
99.	betB	1470	ECOBET
100.	betI	585	ECOBET
101.	betT	2031	ECOBET
102.	fecA	2325	ECOFEC
103.	lacA	609	ECOLACA
104.	lacA	612	ECOLAC
105.	lacY	1254	ECOLAC
106.	lacY	1251	ECOLACY
107.	lacZ	3075	ECOLAC
108.	lacI	1180	ECOLAC
109.	hemB	972	ECOHEMB
110.	hemB	975	ECOHEMBA
111.	sbmA	1221	ECOSBMA
112.	phoA	1416	ECOPHOAAZ, ECOPHOAA, ECOPHOAC
113.	phoA	1416	ECOPHOAD

Table 1 (a) (Continued)

ID	Gene	Length (bases)	GenBank Locus
114.	phoA	1416	ECOPHOAB
115.	phoA	1416	ECOPHOAE
116.	phoA	1416	ECOPHOAF
117.	phoA	1416	ECOPHOAG
118.	phoA	1416	ECOPHOAH
119.	phoA	1416	ECOPHOAI
120.	phoA	1419	ECOPHOAJ
121.	proC	810	ECOPROC
122.	aroL	525	ECOAROLM
123.	aroM	678	ECOAROLM
124.	phoR	1296	ECOPHORG
125.	tgt	1128	ECOTGT
126.	ORF	333	ECOTGT
127.	ORF1	303	ECOSECDF
128.	ORF2	327	ECOSECDF
129.	ORF3	330	ECOSECDF
130.	secD	1845	ECOSECDF
131.	secF	969	ECOSECDF
132.	ORF6	327	ECOSECDF
133.	lon	2352	ECOLON
134.	tesB	861	ECOTESB
135.	nusB	420	ECONUSAA, ECONUSB
136.	apt	552	ECOAPT, ECOAPTADK
137.	dnaX(dnaZX)	1932	ECOAPTADK, ECODNAZX, ECOZXPIII
138.	dnaX(alt.)	1863	ECOZXPIII
139.	ORF12	330	ECOAPTADK, ECODNAZX
140.	recR	606	ECOAPTADK, ECODNAZX
141.	recR	606	ECORECR
142.	hptG(C62.5)	1875	ECOAPTADK, ECODNAZX
143.	adk	645	ECOADK, ECOAPTADK
144.	ushA	1653	ECOUSHA
145.	rhsD	4281	ECORHSDG
146.	purE	510	ECOPUREK
147.	purK	1068	ECOPUREK
148.	cysS	1386	ECOCYSSG
149.	cysS	1386	ECOCYSSGEN
150.	argU(tRNA)	77	ECOINTDLPT, ECOINTDLP
151.	int	1164	ECOINTDLPT, ECOINTDLP
152.	pheP	1377	ECOPHEPA
153.	envY	762	ECOENVY
154.	fepA	2238	ECOFEPAA
155.	fes	1125	ECOFE6
156.	entF	3882	ECOENTF
157.	fepC	816	ECOFEPCDG
158.	fepG	993	ECOFEPCDG
159.	fepD	1005	ECOFEPCDG
160.	fepB	957	ECOFEPB
161.	entC	1176	ECOENTC, ECOENTCE
162.	rna	807	ECORIBI34
163.	rlpA	1089	ECORLPA
164.	leuS	2583	ECOLEUS
165.	asnB	1665	ECOASNB
166.	nagD	750	ECONAGACD
167.	nagC	1218	ECONAGACD
168.	nagA	1146	ECONAGACD
169.	nagB	801	ECONAGBE
170.	nagE	1947	ECONAGBE
171.	glnS	1656	ECOGLNS, ECOGLNSA
172.	glnS	1653	ECOGLNT
173.	kdpE	678	ECOKDPDE
174.	kdpD	2685	ECOKDPDE
175.	kdpC	573	ECOKDPABC
176.	kdpB	2049	ECOKDPABC
177.	kdpA	1674	ECOKDPABC
178.	fhuE	2187	ECOFHUE1
179.	ORF169	510	ECOPHRORF
180.	phr	1419	ECOPHRORF
181.	gltA	1284	ECOGLTA
182.	sdhC	390	ECOGLTA, ECOSDHACD
183.	sdhD	348	ECOGLTA, ECOSDHACD
184.	sdhA	1767	ECOGLTA
185.	sdhB	717	ECOGLTA
186.	sucA	2799	ECOSUCA
187.	sucA	2802	ECOGLTA
188.	sucB	1218	ECOGLTA
189.	sucC	1167	ECOGLTA
190.	sucD	870	ECOGLTA
191.	tolA	1266	ECOTOLAB
192.	tolB	1296	ECOTOLAB
193.	nadA	843	ECONADA
194.	aroG	1053	ECOAROG
195.	galK	1149	ECOGALK
196.	chlD	903	ECOCHLJD
197.	ORF	477	ECOBIO
198.	bioA	1293	ECOBIO
199.	bioB	1041	ECOBIO
200.	bioF	1155	ECOBIO
201.	bioC	756	ECOBIO
202.	bioD	660	ECOBIO
203.	uvrB	2019	ECOUVRB1
204.	uvrB	2022	ECOUVRB
205.	uvrB	2022	ECOUVRB2
206.	glnQ	723	ECOGLNHPQ
207.	glnP	660	ECOGLNHPQ
208.	glnH	747	ECOGLNHPQ
209.	chlE	1236	ECOCHLEN
210.	chlN	750	ECOCHLEN
211.	grx	258	ECOGRX
212.	deoR	759	ECODEOR
213.	serS	1293	ECOSERS
214.	serC	1089	ECOAROA
215.	aroA	1284	ECOAROA
216.	ORF	348	ECORSPA
217.	rpsA	1668	ECORSPA
218.	trxB	966	ECOTRXB
219.	aspC	1191	ECOASPC
220.	ompF	1089	ECOOMPF
221.	asnS	1401	ECOTGASNS
222.	pncB	1203	ECOPNCB
223.	pepN	2613	ECOPEPN
224.	ompA	1041	ECOOMPA
225.	sulA	510	ECOONPA
226.	ndh	1305	ECONDH
227.	rlpA	1089	ECORLPA
228.	appA	1299	ECOAPPAA
229.	putP	1509	ECOPUTP
230.	ORF	558	ECOPYRH
231.	pyrC	1044	ECOPYRH
232.	pyrC	1047	ECOPYRC
233.	ams	2448	ECOAMSG
234.	ptsG	1434	ECOPTSG
235.	g30k	522	ECORPMFA
236.	rpmF	174	ECORPMFA

Table 1 (a) (Continued)

ID	Gene	Length (bases)	GenBank Locus
237.	ORF	879	ECOPIN1
238.	ORF	435	ECOPIN1
239.	pin	555	ECOPIN
240.	pin	552	ECOPIN1
241.	fadR	720	ECOFADR
242.	umuC	1269	ECOUMUCD
243.	umuD	420	ECOUMUCD, ECOUMUDC
244.	hemA	1257	ECOHEMA
245.	hemA	1257	ECOHEMAA
246.	kdsA	855	ECOKDSA
247.	gdhA	1344	ECOGDHA, ECOGDHAK
248.	narL	651	ECONARLX, ECONARXL
249.	narX	1635	ECONARLX
250.	narX	1797	ECONARXL
251.	narK	1392	ECONARK
252.	narH	1536	ECONARGHJ
253.	narJ	711	ECONARI
254.	narI	678	ECONARI
255.	tyrT(α tRNA)	85	ECOTGY1
256.	tyrT(β tRNA)	85	ECOTGY1
257.	tpr	102	ECOTGY1
258.	tdk	615	ECOTDK
259.	tdk	618	ECOTGKG
260.	tonB	735	ECOTONB
261.	P14-protein	399	ECOTONB
262.	trpA	804	ECOTRPX, ECOTRPA
263.	trpA	807	ECOTGP
264.	trpB	1191	ECOTRPX
265.	trpB	1194	ECOTGP
266.	trpC	1350	ECOTRPX
267.	trpC	1359	ECOTGP
268.	trpD	1593	ECOTRPX
269.	trpD	1596	ECOTGP
270.	trpE	1560	ECOTRPX
271.	trpE	1563	ECOTGP
272.	trpL	45	ECOTGP
273.	btuR	591	ECOBTUR
274.	topA	2595	ECOTOPA
275.	cysB	975	ECOCYSB, ECOCYSBA
276.	pyrF	738	ECOPYRF
277.	ORF	327	ECOPYRF
278.	tyrR	1542	ECOTYRR
279.	dcp	2046	ECODCPG
280.	trg	1608	ECOTRG
281.	narZ	3738	ECONARZYW
282.	narY	1542	ECONARZYW
283.	narW	693	ECONARZYW
284.	narV	678	ECONARZYW
285.	relB	240	ECORELB
286.	dicBS	189	ECOICABC
287.	dicBL	333	ECOICABC
288.	dicA	408	ECOICABC
289.	dicC	231	ECOICABC
290.	pntA	1509	ECOPNTAB
291.	pntB	1389	ECOPNTAB
292.	fumC	1404	ECOFUMC
293.	fumA	1647	ECOFUMA
294.	manA	1176	ECOMANAA
295.	malI	977	ECOMALAA
296.	malI	978	ECOMALIXA
297.	malX	1593	ECOMALAA
298.	uidA	1809	ECOUIDAA
299.	add	999	ECOADD
300.	sodB	582	ECOSODB
301.	purR	1023	ECOPURRP
302.	lpp	237	ECOLPP
303.	aroD	723	ECOAROG
304.	aroH	1044	ECOAROH
305.	aroH	1047	ECOAROHA
306.	pps	2379	ECOPPSGEN
307.	pps	2382	ECOPEPSYN
308.	ORF17	465	ECOBTUCED
309.	btuD	750	ECOBTUCED
310.	btuE	552	ECOBTUCED
311.	btuC	981	ECOBTUCED
312.	himA	300	ECOHIMA
313.	himA	300	ECOTHRINF
314.	pheT	2388	ECOHIMA, ECOTHRINF
315.	pheS	996	ECOTHRINF
316.	pheS	984	ECOHIMA
317.	pheS(att.)	45	ECOTHRINF, ECOHIMB
318.	rplT	357	ECOHIMA
319.	rplT	357	ECOTHRINF
320.	infC	543	ECOHIMA
321.	infC	543	ECOTHRINF
322.	thrS	1929	ECOTHRINF
323.	katE	2262	ECOKATE
324.	celF	1116	ECOCELOPE
325.	celF	1119	ECOCELA
326.	celD	840	ECOCELOPE
327.	celD	843	ECOCELA
328.	celC	348	ECOCELOPE
329.	celC	351	ECOCELA
330.	celB	1251	ECOCELOPE
331.	celB	1254	ECOCELA
332.	celA	318	ECOCELOPE
333.	celA	321	ECOCELA
334.	ansA	969	ECOANSORA
335.	ORF	642	ECOANSORA
336.	gap	996	ECOGAP
337.	pabB	1362	ECOPABB
338.	manX	945	ECOMANXY
339.	manX	948	ECOMANXF
340.	zwf	1473	ECOZWF
341.	ruvB	1011	ECORUVAB, ECORUVABA
342.	ruvA	612	ECORUVAB, ECORUVABA
343.	argS	1734	ECOARGS
344.	cheZ	645	ECOCHE3
345.	cheY	390	ECOCHE3
346.	cheY	390	ECOCHEY
347.	cheB	1050	ECOCHE3
348.	tar	1659	ECOTARX
349.	tar	1662	ECOCHE2
350.	cheW	504	ECOCHE1
351.	cheA	1965	ECOCHEA
352.	cheA(alt.)	1674	ECOCHEA
353.	motB	888	ECOMOTAB
354.	glyW(tRNA)	76	ECOPGSA
355.	pgsA	651	ECOPGSA
356.	tyrP	1215	ECOTYRPA
357.	dcm	1419	ECODCM, ECODCMA
358.	ORF2	507	ECODCM
359.	ORF3	471	ECODCM, ECODCMA
360.	ORF4	921	ECODCM

Table 1 (a) (Continued)

ID	Gene	Length (bases)	GenBank Locus
361.	rcsA	624	ECORCSA
362.	amn	1452	ECOAMN
363.	serU(tRNA)	88	ECOTGSSD
364.	sbcB	1401	ECOSBCB
365.	hisL	51	ECOHISOPA
366.	hisG	900	ECOHISOPA
367.	hisD	1309	ECOHISOPA
368.	hisC	1071	ECOHISOPA
369.	hisB	1068	ECOHISOPA
370.	hisH	591	ECOHISOPA
371.	hisA	738	ECOHISOPA
372.	hisF	777	ECOHISOPA
373.	hisI	612	ECOHISOPA
374.	gnd	1407	ECOGND
375.	gnd	1407	ECOGNDG
376.	gnd	1407	ECONDGH
377.	gnd	1407	ECONDGK
378.	gnd	1407	ECONDGL
379.	gnd	1407	ECONDGM
380.	araF	990	ECOARAFGH
381.	araG	1515	ECOARAFGH
382.	araH	990	ECOARAFGH
383.	alkA	849	ECOALKA
384.	mrp	1110	ECOMRPMET
385.	metG	2034	ECOMRPMET
386.	mglB	999	ECOMGLABCO
387.	mglA	1521	ECOMGLABCO
388.	mglC	1011	ECOMGLABCO
389.	dld	1716	ECODLD, ECODLDH
390.	alkB	651	ECOADAB
391.	ada	1065	ECOADA, ECOADAA
392.	ompC	1104	ECOOMPC
393.	rcsB	2802	ECORCSBC
394.	rcsC	651	ECORCSBC
395.	gyrA	2634	ECOGYRAAM
396.	gyrA	2628	ECOGYRA, ECOGYRA1
397.	pufX	720	ECOGYRAM
398.	glpT	1359	ECOGLPT
399.	glpA	1629	ECOGLPA
400.	glpC	1260	ECOGLPA
401.	glpB	1191	ECOGLPA
402.	menD	1389	ECOMEND
403.	ackA	1203	ECOACKA
404.	hisP	774	ECOHISMP
405.	dedF	570	ECOHISPUR2
406.	purF	1515	ECOPURF
407.	ORF(17.9)	489	ECOPURF
408.	hisT	813	ECOHIST1
409.	usg	1014	ECOHIST1
410.	pdxB	1137	ECOPDXB
411.	fabB	1221	ECOFABB
412.	aroC	1071	ECOAROCG, ECOAROCX
413.	aroC	1086	ECOAROC
414.	fadL	1152	ECOFADL
415.	fadL	1347	ECOFADLA
416.	dsdC	750	ECODSDAA
417.	dsdC(alt.)	795	ECODSDAA
418.	ORF	1020	ECODSDAA
419.	dsdA	1329	ECODSDA
420.	valU(α tRNA)	76	ECOVALU
421.	valU(β tRNA)	76	ECOVALU
422.	valU(γ tRNA)	76	ECOVALU
423.	lysV	76	ECOVALU
424.	lig	2016	ECOLIG, ECOLIGA
425.	cysK	972	ECOCYSPTS
426.	ptsH	258	ECOCYSPTS, ECOPTSH, ECPTSHI
427.	ptsI	1728	ECOPTSHI
428.	crr	510	ECOPTSHI
429.	dapE	432	ECOPURCA
430.	purC	714	ECOPURCA
431.	purM	1038	ECOPURMN
432.	purN	639	ECOPURMN
433.	guaA	1578	ECOGUABA
434.	guaB	1467	ECOGUABA
435.	guaB	1533	ECOGUAB
436.	glyA	1251	ECOGLYB
437.	glyA	1254	ECOGLYA
438.	lep	972	ECOLEP
439.	lepA	1797	ECOLEP
440.	recO	729	ECORECO
441.	ung	690	ECOUNG
442.	12785MwProtein	348	ECOPHEAB
443.	pheL	48	ECOPHEAB, ECOPHEAC
444.	pheA	1161	ECOPHEAB
445.	tyrA	1122	ECOPHEAB
446.	aroF	1071	ECOPHEAB
447.	recN	1704	ECORECN
448.	rplS	348	ECOTRMD
449.	trmD	768	ECOTRMD
450.	ORF(21kD)	558	ECOTRMD
451.	rpsP	249	ECOTRMD
452.	grpE	594	ECOGRPE
453.	proV	1203	ECOPROU
454.	proW	1065	ECOPROU
455.	proX	993	ECOPROU
456.	gabT	1281	ECOGABT
457.	recA	1059	ECORECE
458.	recA	1062	ECORECA
459.	mutS	2562	ECOMUTS
460.	katF	1089	ECOKATF
461.	cysC	606	ECOCYSDNC
462.	cysN	1428	ECOCYSDNC
463.	cysD	909	ECOCYSDNC
464.	cysH	732	ECOCYSH
465.	cysH	735	ECOCYSJIHA
466.	cysI	1719	ECOCYSJIHA
467.	cysJ	1800	ECOCYSJIHA
468.	pyrG	1638	ECOPYRG
469.	relA	2235	ECORELA
470.	fucO	1152	ECOFUCAO, ECOFUCOSE
471.	fucA	648	ECOFUCAO, ECOFUCOSE
472.	fucP	1317	ECOFUCOSE
473.	fucI	1776	ECOFUCOSE
474.	fucK	1449	ECOFUCOSE
475.	fucU	402	ECOFUCOSE
476.	fucR	732	ECOFUCOSE
477.	recD	1827	ECORECD
478.	recB	3543	ECORECB
479.	ORF1	804	ECORECC
480.	ORF2	366	ECORECC
481.	ORF3	324	ECORECC
482.	recC	3369	ECORECC
483.	thyA	795	ECOTHYA

Table 1 (a) (Continued)

ID	Gene	Length (bases)	GenBank Locus
484.	galR	1029	ECOGALR
485.	galR	1032	ECOGALLYS
486.	lysA	1263	ECOGALLYS
487.	lysR	963	ECOGALLYS
488.	xprB	897	ECORECJXPR
489.	xprA	708	ECORECJXPR
490.	recJ	1737	ECORECJXPR
491.	ORF5	594	ECOFDAPGK
492.	ORF4	861	ECOFDAPGK
493.	fba	1080	ECOFDAPGK
494.	pgk	1164	ECOFDAPGK
495.	ORF	1020	ECOFDAPGK
496.	ORF3	474	ECOFDAPGK
497.	ORF2	402	ECOFDAPGK
498.	speB	921	ECOSPEAA
499.	speA	1977	ECOSPEA
500.	metK	1155	ECOMETK
501.	pheV(tRNA)	76	ECOPHEV
502.	nupG	1257	ECONUPG
503.	speC	2196	ECOSPEC
504.	exbD	426	ECOEXBBD
505.	exbB	735	ECOEXBBD
506.	metC	1188	ECOMETC
507.	tolC	1470	ECOTOLC
508.	cca	1239	ECOCCA
509.	rpsU	216	ECORPSRPO
510.	dnaG	1740	ECODNAG
511.	dnaG	1746	ECORPSRPO
512.	rpoD	1842	ECORPSRPO
513.	ebgR	984	ECOEBG, ECOEBGRA
514.	ebgA	3096	ECOEBG, ECOEBGRA
515.	ebgC	519	ECOEBG, ECOEBGRA
516.	mtr	1245	ECOMTR, ECOMTRA
517.	pnp	2136	ECORPSOP
518.	rpsO	270	ECORPSO, ECRPSJ, ECORPSOP, ECOSOX, ECORPSPNP
519.	infB(β)	2202	ECONUSA
520.	infB(α)	2673	ECONUSA
521.	nusA	1485	ECONUSA
522.	ORF(15kD)	423	ECONUSA
523.	metY(tRNA)	77	ECONUSA
524.	gltB	4545	ECOGLTB
525.	ssp	639	ECOSSPG
526.	rplM	429	ECORPSI
527.	rpsI	393	ECORPSI
528.	fis	297	ECOFIS, ECOFISA
529.	rrfD(5SrRNA)	120	ECORGNDIS
530.	thrV(tRNA)	76	ECORGNDIS
531.	rrfD(5SrRNA)	120	ECORGNDIS
532.	tag	564	ECOTAG
533.	trkA	1377	ECOTRKAG
534.	rpoA	990	ECORPOA
535.	secY	1332	ECORPLN
536.	rplO	435	ECORPLN
537.	rpmD	180	ECORPLN
538.	rpsE	504	ECORPLN
539.	rplR	354	ECORPLN
540.	rplF	534	ECORPLN
541.	rpsH	393	ECORPLN
542.	rpsN	300	ECORPLN
543.	rplE	540	ECORPLN

ID	Gene	Length (bases)	GenBank Locus
544.	rplX	315	ECORPLN
545.	rplN	372	ECORPLN
546.	rpsQ	255	ECORPOS10
547.	rpmC	192	ECORPOS10
548.	rplP	411	ECORPOS10
549.	rpsC	702	ECORPOS10
550.	rplV	333	ECORPOS10
551.	rpsS	279	ECORPOS10
552.	rplB	822	ECORPOS10
553.	rplW	303	ECORPOS10
554.	rplD	606	ECORPOS10
555.	rplC	630	ECORPOS10
556.	rpsJ	312	ECORPOS10, ECORPSJ
557.	rpsL	375	ECOSTR1
558.	crp	633	ECOCRP
559.	nirB	2517	ECONIRBC
560.	nirC	555	ECONIRBC
561.	ORF190	573	ECOFIC1
562.	ORF1	168	ECOPABAA
563.	fic	600	ECOPABAA
564.	fic	603	ECOFIC1
565.	dam	837	ECODAM
566.	aroB	1089	ECOAROB
567.	envZ	1353	ECOOMPB
568.	ompR	720	ECOOMPB
569.	bioH	771	ECOBIOH
570.	malQ	2085	ECOMALQP
571.	malT	2706	ECOMALT
572.	glpR	900	ECOGLPREG, ECOGLPRZ
573.	glpG	831	ECOGLPREG, ECOGLPRZ
574.	glpE	297	ECOGLPRZ
575.	glpE	396	ECOGLPREG
576.	glpD	1506	ECOSNGLPD
577.	glpD	1512	ECOGLPDA
578.	glgA	1434	ECOGLGA
579.	glgC	1296	ECOGLG
580.	glgX	1524	ECOGLG
581.	glgB	2187	ECOGLGBA
582.	asd	1104	ECOASD
583.	ugpC	1071	ECOUGP
584.	ugpE	846	ECOUGP
585.	ugpA	888	ECOUGP
586.	ugpB	1317	ECOUGP
587.	livF	714	ECOLIVHMGF
588.	livG	768	ECOLIVHMGF
589.	livM	1275	ECOLIVHMGF
590.	livH	927	ECOLIVHMGF
591.	livK	1110	ECOLIVHMGF
592.	livJ	1104	ECOLIVHMGF
593.	ORF(20kD)	597	ECOFTSYEX
594.	ftsY	1494	ECOFTSYEX
595.	ftsE	669	ECOFTSYEX
596.	ftsX	1059	ECOFTSYEX
597.	gor	1353	ECOGOR
598.	bisC	2181	ECOBISCASD
599.	glyS(α)	912	ECOGLYS
600.	glyS(β)	2070	ECOGLYS
601.	xylA	1002	ECOXYLAB
602.	xylA	1323	ECOXYLABA
603.	xylB	1455	ECOXYLABA
604.	ORF	438	ECOXYLABA
605.	selB	1842	ECOSELB

Table 1 (a) (Continued)

ID	Gene	Length (bases)	GenBank Locus
606.	rhsA	4134	ECORHSA
607.	mtlA	1911	ECOMTLA, ECOMTLB
608.	cysE	822	ECOCYSE, ECOCYSXE
609.	cysX	393	ECOCYSXE
610.	kbl	1197	ECOKBLTDH
611.	tdh	1026	ECOKBLTDH
612.	envC	1155	ECOENVCD
613.	envD	2895	ECOENVCD
614.	rpmG	168	ECORPMBG
615.	rpmB	237	ECORPMBG
616.	dut	453	ECODUTPYR
617.	dut	456	ECOPYREA
618.	ORF	633	ECODUTPYR
619.	pyrE	636	ECODUTPYR
620.	uhpT	1392	ECOUHP
621.	uhpC	1323	ECOUHP
622.	uhpB	1503	ECOUHP
623.	uhpA	591	ECOUHP
624.	ilvN	291	ECOILVBPR
625.	ilvB	1689	ECOILVBPR
626.	ilvBN(leader)	99	ECOILVBPR
627.	ORF1	399	ECORECFA
628.	gyrB	2415	ECORECFA
629.	recF	1074	ECORECFA, ECORECF
630.	dnaN	1101	ECODNAAOP
631.	dnaA	1404	ECODNAAOP
632.	ORF	105	ECORPMH
633.	rpmH	138	ECORPMH
634.	rpmH	141	ECODNAAOP, ECORNPA
635.	rnpA	360	ECORNPA
636.	tnaA	1416	ECOTNAA
637.	phoU	726	ECOPHOS, ECOPHOWTU
638.	pstB	774	ECOPHOS ECOPHOWTU
639.	pstA(phoT)	891	ECOPHOS, ECOPHOWTU
640.	pstC	960	ECOPHOS
641.	phoS	1041	ECOPHOS
642.	glmS	1830	ECOUNCC
643.	ORF	1371	ECOUNCC
644.	atpC	420	ECOUNC, ECOUNCC
645.	atpD	1383	ECOUNC, ECOUNCC
646.	atpG	864	ECOUNC, ECOUNCC
647.	atpA	1542	ECOUNC, ECOUNCC
648.	atpH	531	ECOATPXB
649.	atpH	534	ECOUNC, ECOUNCC
650.	atpF	468	ECOATPXB
651.	atpF	471	ECOUNC, ECOUNCC
652.	atpE	237	ECOATPXB
653.	atpE	240	ECOUNC, ECOUNCC
654.	atpB	603	ECOATPXB
655.	atpB	813	ECOATPXB
656.	atpB	816	ECOUNC, ECOUNCC
657.	atpI	393	ECOUNC, ECOUNCC
658.	gidB	624	ECOUNCC
659.	gidA	1887	ECOUNCC
660.	ORF(16kD)	444	ECOORI
661.	asnA	990	ECOASNA
662.	asnA	993	ECOORI
663.	rbsD	420	ECORBS
664.	rbsA	1506	ECORBS
665.	rbsC	966	ECORBS
666.	rbsB	891	ECORBS
667.	rbsK	930	ECORBS
668.	trpT(tRNA)	76	ECOTGWS
669.	kdsB	747	ECOKDSB
670.	ilvL	99	ECOILVGMED
671.	ilvL	99	ECOILVX
672.	ilvG	903	ECOILVX
673.	ilvG	1647	ECOILVGMED
674.	ilvG	1647	ECOILVGE
675.	ilvM	261	ECOILVGE
676.	ilvE	930	ECOILVGMED, ECOILVGE
677.	ilvD	1848	ECOILVGMED
678.	ilvD	1851	ECOILVGE
679.	ilvA	1545	ECOILVGMED, ECOILVGE
680.	ilvY	894	ECOILVYC
681.	ilvC	1476	ECOILVYC
682.	trxA	330	ECORH0A, ECOTRXA
683.	trxA	384	ECORH0B
684.	dapF	828	ECODAPF
685.	hemC	942	ECOHEMC, ECOHEMCD
686.	hemD	741	ECOHEMCD
687.	ORFX	1182	ECOHEMCD
688.	ORFY	1197	ECOHEMCD
689.	uvrD	2163	ECOUVRD, ECOUVRD02
690.	recQ	1833	ECORECQ
691.	fadA	1161	ECOFADAC
692.	fadA	1164	ECOFADAB, ECOFADBA
693.	fadB	2187	ECOFADAC
694.	fadB	2190	ECOFADAB, ECOFADBA
695.	ORF315	945	ECOFADAC
696.	glnG	1407	ECOGLN
697.	glnL	1050	ECOGLN
698.	glnA	1410	ECOGLNAB
699.	cpxA	1367	ECOCPXA
700.	cpxA	1374	ECOCPXA
701.	pfkA	1023	ECOCDHA
702.	sbp	990	ECOCDHA
703.	cdh	756	ECOCDH
704.	cdh	753	ECOCDHA
705.	glpF	846	ECOGLPFK, ECOGLPF
706.	glpK	1509	ECOGLYK
707.	metF	891	ECOMETF
708.	katG	2181	ECOKATGA
709.	ppc	2652	ECOPPCG
710.	argC	1005	ECOARGBCH
711.	argB	777	ECOARGBCH
712.	btuB	1845	ECOBTUB
713.	rrsB(16SrRNA)	1542	ECORGNB
714.	gltT(tRNA)	76	ECORGNB
715.	rrlB(23SrRNA)	2904	ECORGNB
716.	rrfB(5SrRNA)	120	ECORGNB
717.	ORF	1029	ECOBIRA
718.	birA	966	ECOBIRA
719.	rts	1011	ECORTSA
720.	thrU(tRNA)	76	ECOTGTUFB, ECOTUFB1
721.	tyrU(tRNA)	85	ECOTGTUFB, ECOTUFB1
722.	glyT(tRNA)	75	ECOTGTUFB, ECOTUFB1
723.	thrT(tRNA)	74	ECOTUFB1
724.	thrT(tRNA)	76	ECOTGTUFB
725.	tufB	1185	ECOTGTUFB
726.	ORF1	414	ECOTGTUFB
727.	ORF2	366	ECOTGTUFB
728.	ORF3	306	ECOTGTUFB
729.	rplK	426	ECORPOBC

Table 1 (a) (Continued)

ID	Gene	Length (bases)	GenBank Locus
730.	rplK	429	ECORPLRPO
731.	rplA	702	ECORPOBC
732.	rplA	705	ECORPLRPO
733.	rplJ	495	ECORPOBC
734.	rplJ	498	ECORPLRPO
735.	rplL	363	ECORPOBC
736.	rplL	366	ECORPLRPO
737.	rpoB	4026	ECORPOBC
738.	rpoB	4029	ECORPLRPO
739.	rpoC	4221	ECORPOBC
740.	rpoC	4224	ECORPLRPO
741.	purD	1287	ECOPURHE
742.	purD	1290	ECOPURHD
743.	purH	1587	ECOPURHE
744.	purH	1590	ECOPURHD
745.	xylE	1476	ECOXYLE
746.	malG	891	ECOMALG
747.	malE	1191	ECOMALB
748.	malK	1113	ECOMALB
749.	lamB	1338	ECOLAMBA
750.	lamB	1341	ECOLAMBBB, ECOMALB
751.	molA	396	ECOMALB
752.	lexA	609	ECOLEXA
753.	plsB	2424	ECOPLSB
754.	dnaB	1416	ECODNAB
755.	tyrB	1194	ECOTYRB, ECOTYRBA
756.	uvrA	2823	ECOUVRAA
757.	ORF742	2229	ECOPHNAQ
758.	phnA	336	ECOPHNAQ
759.	phnB	444	ECOPHNAQ
760.	phnC	789	ECOPHNAQ, ECOPHN
761.	phnD	1017	ECOPHNAQ
762.	phnD	1017	ECOPHN
763.	phnE	831	ECOPHNAQ
764.	phnE1	570	ECOPHN
765.	phnE2	219	ECOPHN
766.	ORF146	441	ECOPHNAQ
767.	phnF	726	ECOPHNAQ, ECOPHN
768.	phnG	453	ECOPHNAQ, ECOPHN
769.	phnH	585	ECOPHNAQ, ECOPHN
770.	ORF269	810	ECOPHNAQ
771.	phnI	1065	ECOPHNAQ, ECOPHN
772.	phnJ	846	ECOPHNAQ, ECOPHN
773.	phnK	759	ECOPHNAQ, ECOPHN
774.	phnL	681	ECOPHNAQ, ECOPHN
775.	ORF114	345	ECOPHNAQ
776.	phnM	1137	ECOPHNAQ, ECOPHN
777.	phnN	558	ECOPHNAQ, ECOPHN
778.	phnO	435	ECOPHNAQ, ECOPHN
779.	ORF126	381	ECOPHNAQ
780.	phnP	759	ECOPHNAQ, ECOPHN
781.	phnQ	408	ECOPHN
782.	phnQ	366	ECOPHNAQ
783.	melA	1356	ECOMELA
784.	melB	1410	ECOMELB
785.	lysU	1506	ECOLYSU
786.	pheU(tRNA)	76	ECOTGF
787.	aspA	1437	ECOASPAG, ECOASPAW
788.	ampC	1134	ECOAMPCFR
789.	frdD	360	ECOAMPCFR
790.	frdC	396	ECOAMPCFR
791.	frdB	735	ECOAMPCFR
792.	frdA	1809	ECOAMPCFR
793.	psd	969	ECOPSD
794.	miaA	951	ECOMIAA
795.	rpsF	396	ECORPSFRI
796.	ORF	315	ECORPSFRI
797.	rpsR	228	ECORPSFRI
798.	rplI	450	ECORPSFRI
799.	cpdB	1941	ECOCPDB
800.	pyrBI(leader)	135	ECOPYRBIB, ECOPYRBIC
801.	valS	2856	ECOVALS
802.	fimB	603	ECOFIMBE
803.	fimE	597	ECOFIMBE
804.	fimA	546	ECOFIMA
805.	fimG	504	ECOFIMFGH
806.	fimH	903	ECOFIMFGH
807.	mcrB	1398	ECOMCRBC
808.	mcrC	1077	ECOMCRBC
809.	hsdR	3102	ECOHSDMSR
810.	hsdS	1233	ECOHSDMSR
811.	hsdS	1332	ECOHSDD
812.	hsdS	1335	ECOHSDSD
813.	hsdS	1392	ECOHSDK
814.	hsdS	1422	ECOHSDB
815.	hsdS	1425	ECOHSDSB
816.	hsdS	1770	ECOHSDSA
817.	hsdS	1785	ECOHSDS
818.	hsdM	1563	ECOHSDMSR
819.	dnaT	540	ECODNATC
820.	dnaC	738	ECODNATC
821.	p-18 protein	498	ECODNATC
822.	rimI	486	ECORIMI
823.	leuV(α tRNA)	87	ECOTRNA
824.	leuV(β tRNA)	87	ECOTRNA
825.	leuV(γ tRNA)	88	ECOTRNA
826.	deoC	777	ECODEOCA1
827.	serB	969	ECOSERB
828.	ORF(SMP)	645	ECOSERB
829.	trpR	327	ECOTRPR
830.	ORF173	414	ECOTRPR
831.	ORF121	366	ECOTRPR
832.	ORF101	306	ECOTRPR
833.	ORF(17kD)	474	ECOPHOM
834.	ORF(28kD)	690	ECOPHOM
835.	phoM	1425	ECOPHOM
836.	cet	1353	ECOCET
837.	ORF83	252	ECOILER
838.	ileR	303	ECOILER

Table 1 (b) Bacteriophage ΦX174

ID	Gene	Length (bases)	GenBank Locus
1.	A	1542	PX1CG
2.	A*	1026	PX1CG
3.	B	363	PX1CG
4.	K	171	PX1CG
5.	C	261	PX1CG
6.	D	459	PX1CG
7.	E	276	PX1CG
8.	J	117	PX1CG
9.	F	1284	PX1CG
10.	G	528	PX1CG
11.	H	987	PX1CG

Table 1 (c) Bacteriophage G 4

ID	Gene	Length (bases)	GenBank Locus
1.	A	1665	PG4CG
2.	A*	1026	PG4CG
3.	B	363	PG4CG
4.	K	171	PG4CG
5.	C	255	PG4CG
6.	D	459	PG4CG
7.	E	291	PG4CG
8.	J	78	PG4CG
9.	F	1284	PG4CG
10.	G	534	PG4CG
11.	H	1014	PG4CG

Table 1 (d) Bacteriophage M13

ID	Gene	Length (bases)	GenBank Locus
1.	II	1233	M13
2.	X	336	M13
3.	V	264	M13
4.	VII	102	M13
5.	IX	99	M13
6.	VIII	222	M13
7.	III	1275	M13
8.	VI	339	M13
9.	I	1047	M13
10.	IV	1281	M13

Table 1 (e) Bacteriophage fd

ID	Gene	Length (bases)	GenBank Locus
1.	II	1233	PFDCG
2.	X	336	PFDCG
3.	V	264	PFDCG
4.	VII	102	PFDCG
5.	IX	99	PFDCG
6.	VIII	222	PFDCG
7.	III	1275	PFDCG
8.	VI	339	PFDCG
9.	I	1047	PFDCG
10.	IV	1281	PFDCG

Table 1 (f) Bacteriophage f1

ID	Gene	Length (bases)	GenBank Locus
1.	II	1233	F1CCG, F1CCGA
2.	X	336	F1CCG, F1CCGA
3.	V	264	F1CCG, F1CCGA
4.	VII	102	F1CCG, F1CCGA
5.	IX	99	F1CCG, F1CCGA
6.	VIII	222	F1CCG, F1CCGA
7.	III	1275	F1CCG, F1CCGA
8.	VI	339	F1CCG, F1CCGA
9.	I	1047	F1CCG, F1CCGA
10.	IV	1281	F1CCG, F1CCGA

Table 1 (g) Bacteriophage Ike

ID	Gene	Length (bases)	GenBank Locus
1.	II	1266	IKECG
2.	X	366	IKECG
3.	V	267	IKECG
4.	VII	99	IKECG
5.	IX	102	IKECG
6.	VIII	249	IKECG
7.	III	1305	IKECG
8.	VI	351	IKECG
9.	I	1098	IKECG
10.	IV	1314	IKECG

Table 1 (h) Bacteriophage P 4

ID	Gene	Length (bases)	GenBank Locus
1.	gop	399	PP4P4CG
2.	beta	1065	PP4P4CG
3.	cII	792	PP4P4CG
4.	int	1320	PP4P4CG
5.	alpha	2311	PP4P4CG
6.	ORF106	318	PP4P4CG
7.	ORF151	453	PP4P4CG
8.	epsilon	285	PP4P4CG
9.	cI	411	PP4P4CG
10.	ORF88	264	PP4P4CG
11.	sid	732	PP4P4CG
12.	delta	498	PP4P4CG
13.	psu	570	PP4P4CG

Table 1 (i) Bacteriophage T 7

ID	Gene	Length (bases)	GenBank Locus
1.	0.3	354	PT7CG
2.	0.4	156	PT7CG
3.	0.5	144	PT7CG
4.	0.6a	164	PT7CG
5.	0.7	1080	PT7CG
6.	1	2652	PT7CG
7.	1.1	129	PT7CG
8.	1.2	258	PT7CG
9.	1.3	1080	PT7CG
10.	1.4	156	PT7CG
11.	1.5	90	PT7CG
12.	1.6	261	PT7CG
13.	1.7	591	PT7CG
14.	1.8	147	PT7CG
15.	2	195	PT7CG
16.	2.5	699	PT7CG
17.	2.8	420	PT7CG
18.	3	450	PT7CG
19.	3.5	456	PT7CG
20.	3.8	366	PT7CG
21.	4a	1701	PT7CG
22.	4.1	123	PT7CG
23.	4b	1512	PT7CG
24.	4.2	339	PT7CG
25.	4.3	213	PT7CG
26.	4.5	270	PT7CG
27.	4.7	408	PT7CG
28.	5	2115	PT7CG
29.	5.3	357	PT7CG
30.	5.5	300	PT7CG
31.	5.7	210	PT7CG
32.	6	1047	PT7CG
33.	6.3	114	PT7CG
34.	6.5	255	PT7CG
35.	6.7	267	PT7CG
36.	7	402	PT7CG
37.	7.3	300	PT7CG
38.	7.7	393	PT7CG
39.	8	1611	PT7CG
40.	9	924	PT7CG
41.	10a	1038	PT7CG
42.	11	591	PT7CG
43.	12	2385	PT7CG
44.	13	417	PT7CG
45.	14	591	PT7CG
46.	15	2244	PT7CG
47.	16	3957	PT7CG
48.	17	1662	PT7CG
49.	17.5	204	PT7CG
50.	18	270	PT7CG
51.	18.5	432	PT7CG
52.	18.7	252	PT7CG
53.	19	1761	PT7CG
54.	19.2	258	PT7CG
55.	19.3	174	PT7CG
56.	19.5	150	PT7CG

Table 1 (j) Bacteriophageλ

ID	Gene	Length (bases)	GenBank Locus
1.	nul	546	LAMCG
2.	A	1926	LAMCG
3.	W	207	LAMCG
4.	B	1602	LAMCG
5.	C	1320	LAMCG
6.	nu3	606	LAMCG
7.	D	333	LAMCG
8.	E	1026	LAMCG
9.	Fi	399	LAMCG
10.	Fii	354	LAMCG
11.	Z	579	LAMCG
12.	U	396	LAMCG
13.	V	741	LAMCG
14.	G	423	LAMCG
15.	T	435	LAMCG
16.	H	2562	LAMCG
17.	M	330	LAMCG
18.	L	699	LAMCG
19.	K	600	LAMCG
20.	I	672	LAMCG
21.	J	3399	LAMCG
22.	lom	579	LAMCG
23.	ORF401	396	LAMCG
24.	ORF314	741	LAMCG
25.	ORF194	423	LAMCG
26.	ea47	435	LAMCG
27.	ea31	2562	LAMCG
28.	ea59	1578	LAMCG
29.	int	1071	LAMCG
30.	xis	219	LAMCG
31.	ea8.5	282	LAMCG
32.	ea22	549	LAMCG
33.	exo	681	LAMCG
34.	bet	786	LAMCG
35.	gam	417	LAMCG
36.	kil	144	LAMCG
37.	cIII	165	LAMCG
38.	ea10	369	LAMCG
39.	ral	201	LAMCG
40.	N	402	LAMCG
41.	rexb	435	LAMCG
42.	rexa	840	LAMCG
43.	cI	714	LAMCG
44.	cro	201	LAMCG
45.	cII	294	LAMCG
46.	O	900	LAMCG
47.	P	702	LAMCG
48.	ren	291	LAMCG
49.	Nin146	441	LAMCG
50.	Nin290	873	LAMCG
51.	Nin57	174	LAMCG
52.	Nin60	183	LAMCG
53.	Nin56	171	LAMCG
54.	Nin204	615	LAMCG
55.	Nin68	207	LAMCG
56.	Nin221	666	LAMCG
57.	Q	624	LAMCG
58.	ORF64	195	LAMCG
59.	S	324	LAMCG
60.	R	477	LAMCG
61.	Rz	642	LAMCG

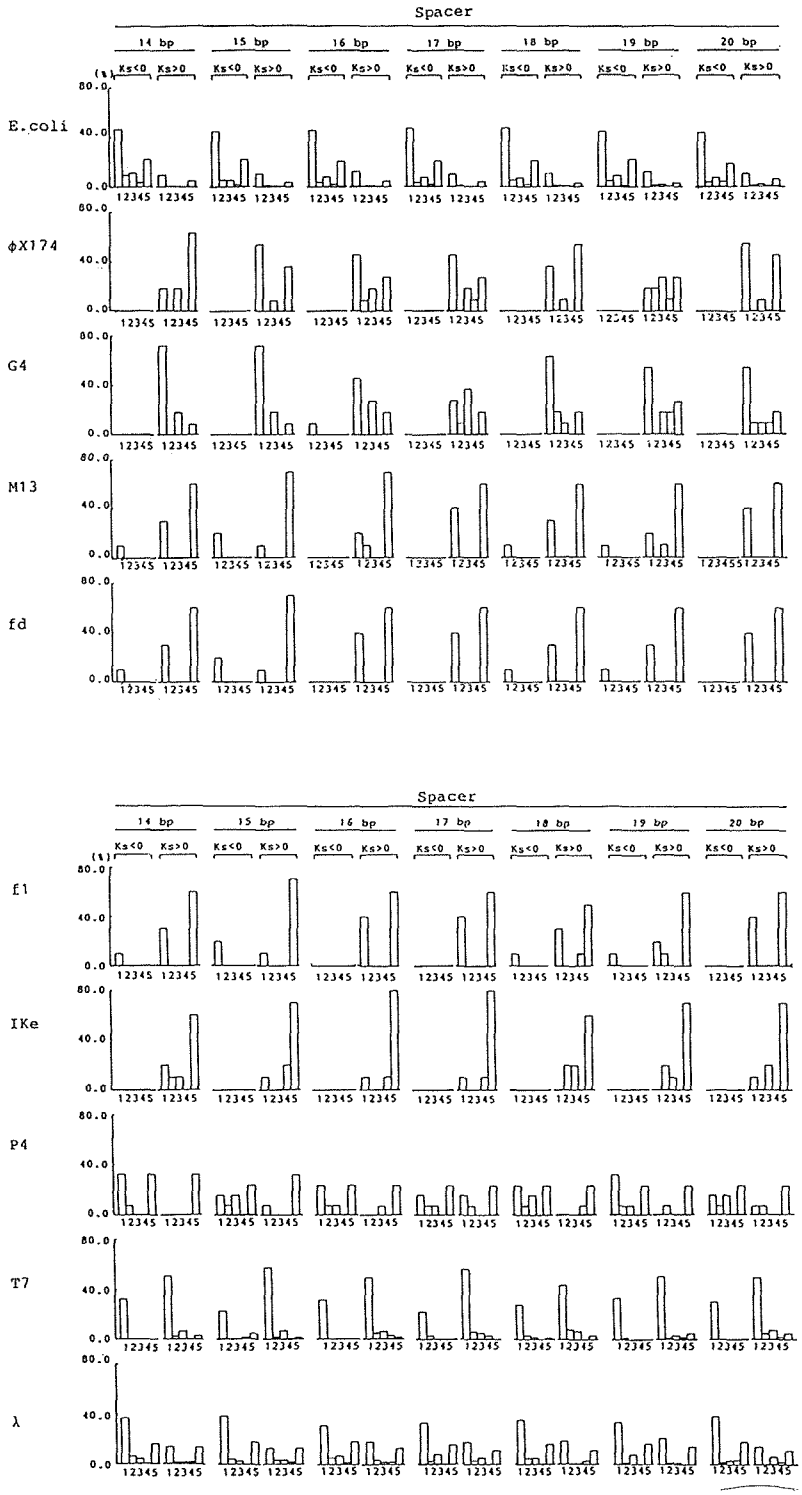
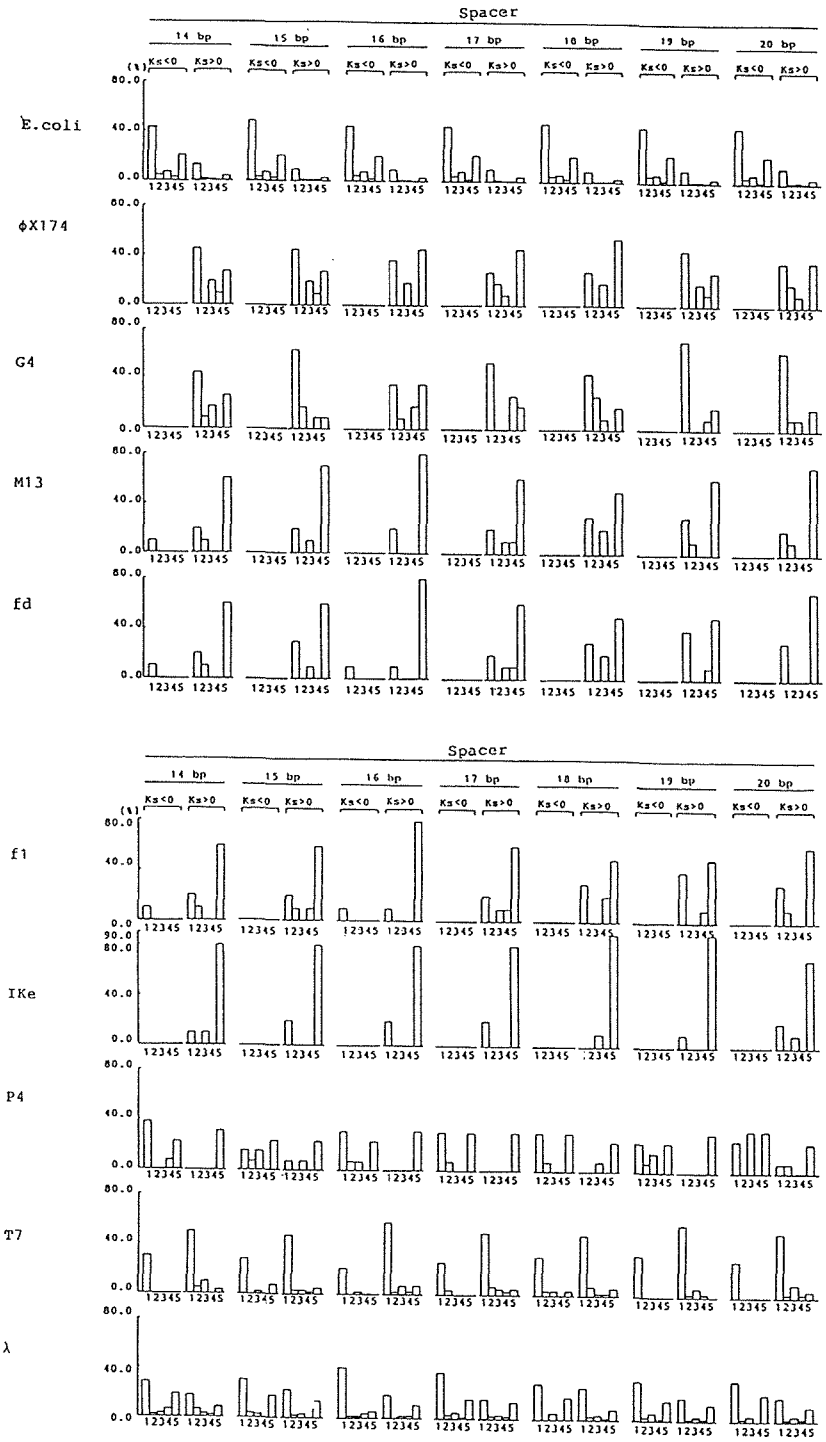
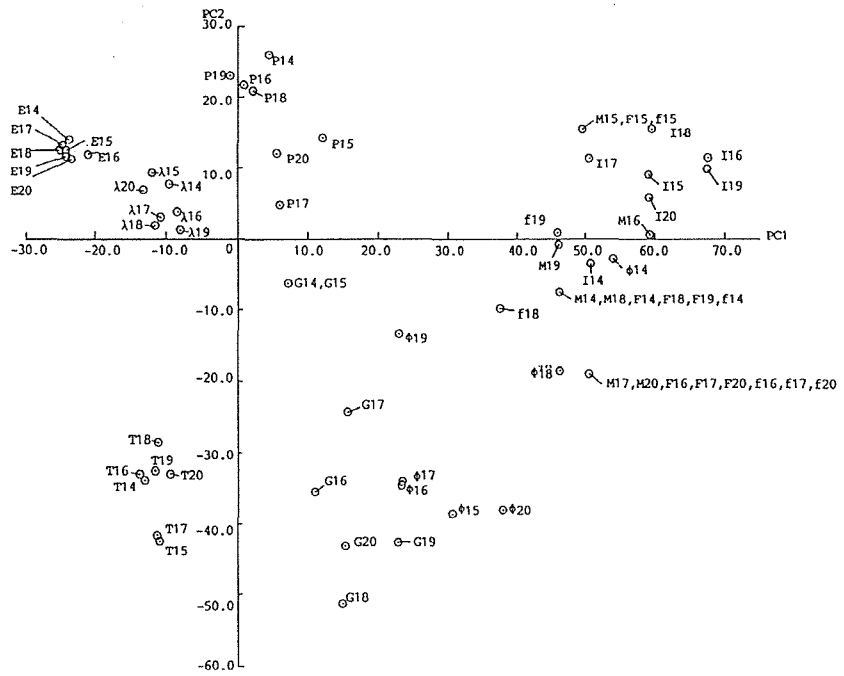


Figure 1. (a) On TS.

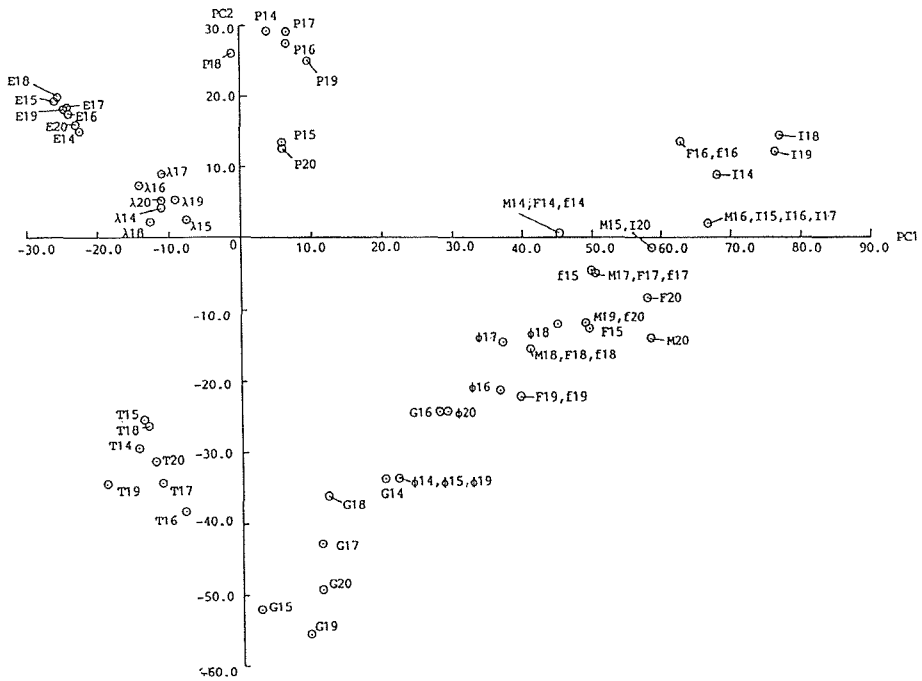


(b) On NTS.

Figure 1 (b) Distributions of percent gene frequencies on TS.
(Levels of significance : 1, $0.05 \leqq p$; 2, $0.025 \leqq p < 0.05$; 3, $0.005 \leqq p < 0.025$; 4, $0.0025 \leqq p < 0.005$; 5, $p < 0.0025$)

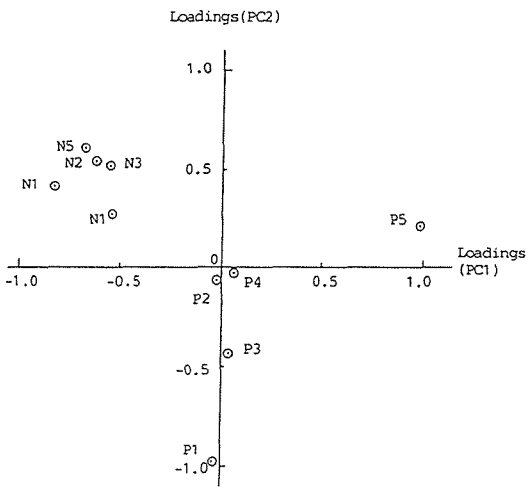


(a) In NTS.

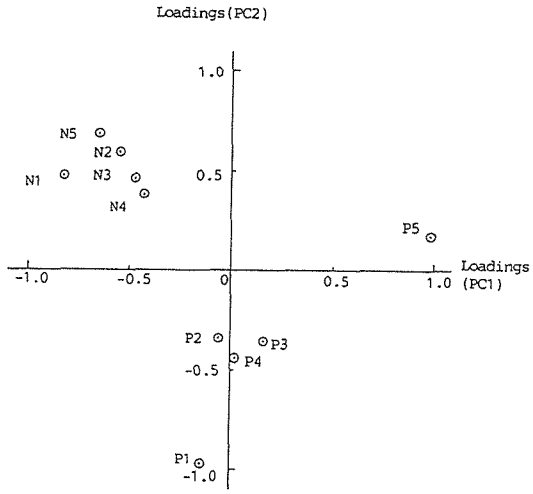


(b) In TS.

Figure 3. Projections by the first two components (E, E. coli; P, P4; λ, lambd da; T, T7; G, G4; F, fd; f, f1; M, M13; I, IKe; Φ. ΦX174; Numerals 14 to 20 represent spacer types.)



(a) On TS.



(b) On NTS.

Figure 2. Factor loadings of the first two principal components (N, Ks< 0 ; P, Ks> 0 ; Level of significance : See Figure 1.)

4. Discussion

The previous paper[8] (A) reported on similar repressibilities, among E. coli and the two temperate phages (P 4 and λ), of occurrences of the promoter-like sequences in coding regions and (B) strongly suggested that supporting much the same repressions of occurrences of the promoter-like sequences in coding regions as E. coli was one of the most important factors required for temperate phages to lysogenize.

Results of PCA (Figure 2 and 3) not only confirm furthermore the previous suggestion[8], but also represent a subtle distinction of the gene distributions among E. coli and its bacteriophages as follows : 〔1〕 phage λ is more similar to E. coli than phage P 4, and 〔2〕 the occurrences of the promoter-like sequences in the coding regions for filamentous phages (M13, fd, f1, and IKe) are more promotive than those for T 7, G 4 and ΦX174. And these results demonstrate that the distribution of gene frequencies by the index, Ks, and statistical significance indicate a genomic regulation in transcription initiation. Besides the previous assertion that various kinds of regulation types of occurrences of the promoter-like sequences are significantly related to polymorphism of transcription units such as polycistronic and monocistronic transcriptions[5-7] and to the inherent life cycles in becteriophages[8], we would like to propose that the strict repression of the promoter-like sequences in E. coli is a fundamental factor to effectively coordinate a web of gene control circuit. Investigation is currently in progress to clarify repressibility of the other consensus sequences of the other σ-factors[15,16].

References

1) Ulrich Siebenlist, Robert B. Simpson and Walter Gibelt: Cell **20**, 269(1980) .
2) Diane K., Howley and William R.McClure: Nucleic Acids Res., **11**, 2237(1983) .
3) Calvin B.Harley and Robert P.Reynolds : Nucleic Acids Res., **15**, 2343(1987) .
4) Martin Rosenverg and Donald Court : Am Rev. Genet., **13**, 319(1979) .
5) Shigehiko Kanaya and Yoshihiro Kudo, Nucleic Acids Symposium Series No.25, 41(1991) .
6) Shigehiko Kanaya and Yoshihiro Kudo, Bull. Yamagata Univ.,(Eng.)**22**, 65(1992).
7) Shigehiko Kanaya and Yoshihiro Kudo, Anal. Sci. **7**, Supple. 739(1991) .
8) Shigehiko Kanaya and Yoshihiro Kudo, Proccedings of the Second International Conference on CAMSE, in press.
9) D.L.Daniels and F.R.Blattner, Nature, **325**, 831(1987) .
10) Gail E.Christie, Annu. Rev. Genet., **24**, 465(1990) .
11) Ira Herskowitz and David Hageu, Annu. Rev. Genet., **14**, 399(1980) .
12) Ben P. H. Peeters, Pik M. Perers, JohnG.G.Schoemakers and Rund N.H.Konings, J. Mol. Biol., **181**, 27(1985).
13) Peter M. G. F. van Wezenkbeek, Theo J.M.Hulsebos, Gene, **16**, 35(1981) .
14) Ewald Beck and Barbara Zink, Gene **16**, 35(1981) .
15) John D.Helmann and Michael J.H., Annu. Rev. Biochem., **57**, 839(1988) .
16) William S. Renikiff, Deborah A. Segele, Deborah W.Cowing, and Carol A.Gross, Annu. Rev. Genet., **19**, 335(1975) .